Diatom.assembly

Fernanda Gonzalez-Saldias

2025-07-01

Contents

1 Set working space

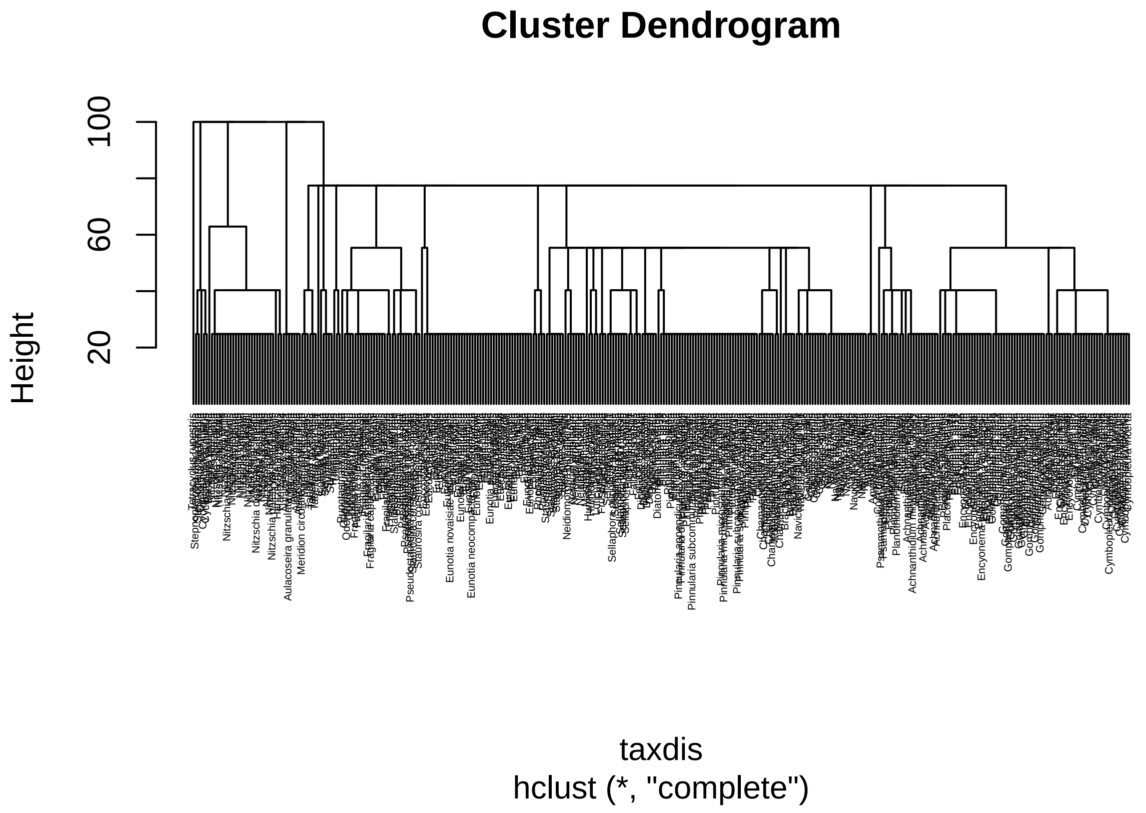
1.1 Load libraries

```
rm(list=ls())
library(vegan)
library(ggplot2)
library(dplyr)
library(readxl)
library(xlsx)
library(cowplot)
library(ape)
library(phyloseq)
library(stringr)
library(iCAMP)
library(NST)
library(dplyr)
library(tidyr)
library(grid)
library(tidyverse)
library(ggvenn)
library("ggVennDiagram")
library(pheatmap)
library(cowplot)
library("pgirmess")
library(ggsignif)
```

2 Community assembly processes

2.1 Load traxonomic tree dataset and deterministic processes (bNTI)

```
taxon_name=data.frame(read_xlsx("data\\Table S2.xlsx",sheet="Taxonomic tree",na=""))
rownames(taxon_name)=taxon_name$Species
Abun_rel=data.frame(read_xlsx("data\\Table S2.xlsx",sheet="Relative abundance",na=""))
comm<-t(sqrt(Abun_rel[8:360]))
TaxonID<-row.names(comm)

# Taxonomic distances from a classification table with variable step lengths.
taxdis <- taxa2dist(taxon_name, varstep=TRUE)
plot(hclust(taxdis), hang = -1, cex=0.3)
```



```
dist.df=data.frame(as.matrix(taxdis))
```

```
# bNTI calculation
#row.names(dist.df)<-names(dist.df)
#set.seed(1)
#NTI.out = bNTIn.p(t(comm), dist.df, nworker = 4, memo.size.GB = 16, weighted = TRUE,
#                   exclude.consp = FALSE, rand = 1000, output.bMNTD = TRUE,
#                    sig.index = "bNTI",
#                   unit.sum = NULL, correct.special = FALSE, detail.null = FALSE,
#                   special.method = "MNTD")

#save(file="bNTI_bioocult_20241202.RData",NTI.out)
```

2.2 Load relative abundance dataset and Raup crick (RC) calculation

```
load("data\\bNTI_bioocult_20241202.RData")
metadata=data.frame(read_xlsx("data\\Table S1.xlsx",sheet="Metadata",na=""))
Abun_rel=data.frame(read_xlsx("data\\Table S2.xlsx",sheet="Relative abundance",na=""))

rownames(Abun_rel)=Abun_rel$Station
rownames(metadata)=metadata$Station

xy <- t(combn(colnames(NTI.out$index), 2))
tmp = data.frame(xy, BNTI = NTI.out$index[xy])
tmp_merge=merge(tmp,metadata[,1:4],by.x="X1",by.y="Station")
indice=merge(tmp_merge,metadata[,1:4],by.x="X2",by.y="Station")
```

```
indice=indice[indice$Local.x==indice$Local.y,]
indice$proc="Stochastic"
indice$proc[indice$BNTI< (-2)]="Homogeneous selection"
indice$proc[indice$BNTI>2]="Heterogeneous selection"
indice$index=paste0(indice$X2,".",indice$X1)
```

3 Mire inside analysis (Local scale)

3.1 RC mire inside

```
comm<-t(sqrt(Abun_rel[8:360]))
TaxonID<-row.names(comm)
meta.groupi<-data.frame(cluster=Abun_rel[,c("Local")])
row.names(meta.groupi)<-Abun_rel$Station


### taxonomic normalized stochasticity ratio
set.seed(1)
tnst=tNST(comm=t(comm),meta.com=NULL,group=meta.groupi,
          dist.method="bray", abundance.weighted=TRUE, rand=1000,
          output.rand=TRUE, nworker=8, LB=FALSE, null.model="PF",
          between.group=TRUE, SES=TRUE, RC=TRUE)

#to extract only the comparisons within each mire
nst.temp<-tnst[["index.pair.grp"]]
nst.temp$index=paste0(nst.temp$name1,".",nst.temp$name2)

#RC Percentage
nst.temp$category<-"bRC<|0.95|"
nst.temp$category[nst.temp$RC.ij.bray>0.95]<-"bRC>0.95"
nst.temp$category[nst.temp$RC.ij.bray< (-0.95)]<-"bRC<-0.95"

#To extract stochastic data
stoch_cate=nst.temp[,c(2,3,10,11,12)]
stoch_cate$proc<-"Ecological drift"
stoch_cate$proc[stoch_cate$RC.ij.bray>0.95]<-"Dispersal limitation"
stoch_cate$proc[stoch_cate$RC.ij.bray< (-0.95)]<-"Homogenizing dispersal"

procesos=merge(indice,stoch_cate,by="index")

procesos$process<-NA
procesos$process=procesos$proc.x
procesos$process=ifelse(procesos$process=="Stochastic",procesos$proc.y,procesos$proc.x)

agg.bRC<-aggregate(RC.ij.bray~process+Local.x,data=procesos,length)
agg.bRC<-agg.bRC%>% complete(process,Local.x)#complete when there is nothing
agg.bRC$RC.ij.bray[is.na(agg.bRC$RC.ij.bray)]<-0#if NA replace with 0

agg.bRC$category<-as.factor(agg.bRC$process) #Set it as factor

#Percentages
gg.tmp<-aggregate(RC.ij.bray~Local.x,data=procesos,length)
gg.bRC_percentage<-merge(agg.bRC,gg.tmp,by="Local.x")

gg.bRC_percentage$perc.RC<-gg.bRC_percentage$RC.ij.bray.x/gg.bRC_percentage$RC.ij.bray.y
```

3.2 Displey panel Figure S3 A

```
cbp1 <- c( "#bdbdbd","#FFDB6D","#1c9099","#fd8d3c","#df65b0"
           ,"#56B4E9","#009E73","#293352","#F0E442","#ff0000")

#define factors and added nationals parks in each mire
agg.bRC_all=merge(agg.bRC,unique(procesos[,6:7]),by = "Local.x")

agg.bRC_all$group<-factor(agg.bRC_all$Local.x,
                          c("PCCO","PCER","PLIO","PRON","PSVB",
                            "OBDC","OBUS","OOTA","OOTS",
                            "GBAD","GMHO","GNAJ","GRAS",
                            "NCLV","NCPG","NDAL","NDCM","NTCC"))
agg.bRC_all$proc<-factor(agg.bRC_all$process,
                         c( "Ecological drift","Dispersal limitation",
                            "Homogenizing dispersal",
                            "Heterogeneous selection","Homogeneous selection"))

agg.bRC_all$Landscape.x<-factor(agg.bRC_all$Landscape.x,
                                    c("Picos de Europa","Ordesa y Monte Perdido",
                                      "Guadarrama","Sierra Nevada"))
prop.bRC<-agg.bRC_all%>%
  ggplot(aes(x=group, y = RC.ij.bray, fill = proc)) +
  geom_col(position = "fill")+
  scale_fill_manual(values = c(cbp1[1:6]),name=" ")+
  labs(title="",x=NULL,y="Community assembly processes (%) ")+
  theme_bw()+
  theme(panel.grid.minor = element_blank(),panel.grid.major = element_blank(),
        axis.text.y = element_text(size=10, color="black",hjust=1,angle = 0),
        axis.text.x = element_blank(),axis.title.x = element_blank(),
        text = element_text(family="serif"))+
  facet_wrap(~Landscape.x,scales = "free_x",nrow = 1)+
  theme(plot.margin = unit(c(0.1,0.15, 0.1,0), "cm"))+
  theme(legend.key.size = unit(0.7, "lines"),
        legend.margin=margin(0,1,0,1),legend.box.margin=margin(-10,-10,-10,-10),
        legend.position = "right",legend.box = "horizontal",
        legend.key = element_rect(color = NA, fill = NA),
        legend.spacing.x = unit(0.2,"cm"),legend.text=element_text(size=8),
        legend.title=element_text(size=6))+
  theme(panel.spacing = unit(0.1, "cm"),strip.text.x = element_blank(),
        strip.text.y = element_text(size = 6, color = "black"),
        panel.border = element_rect(color = "grey", fill = NA))+
  guides(fill=guide_legend(ncol = 2))
legenda_prop.bRC=get_plot_component(prop.bRC, 'guide-box-right', return_all = TRUE)
prop.bRC=prop.bRC+theme(legend.position = "none")
```

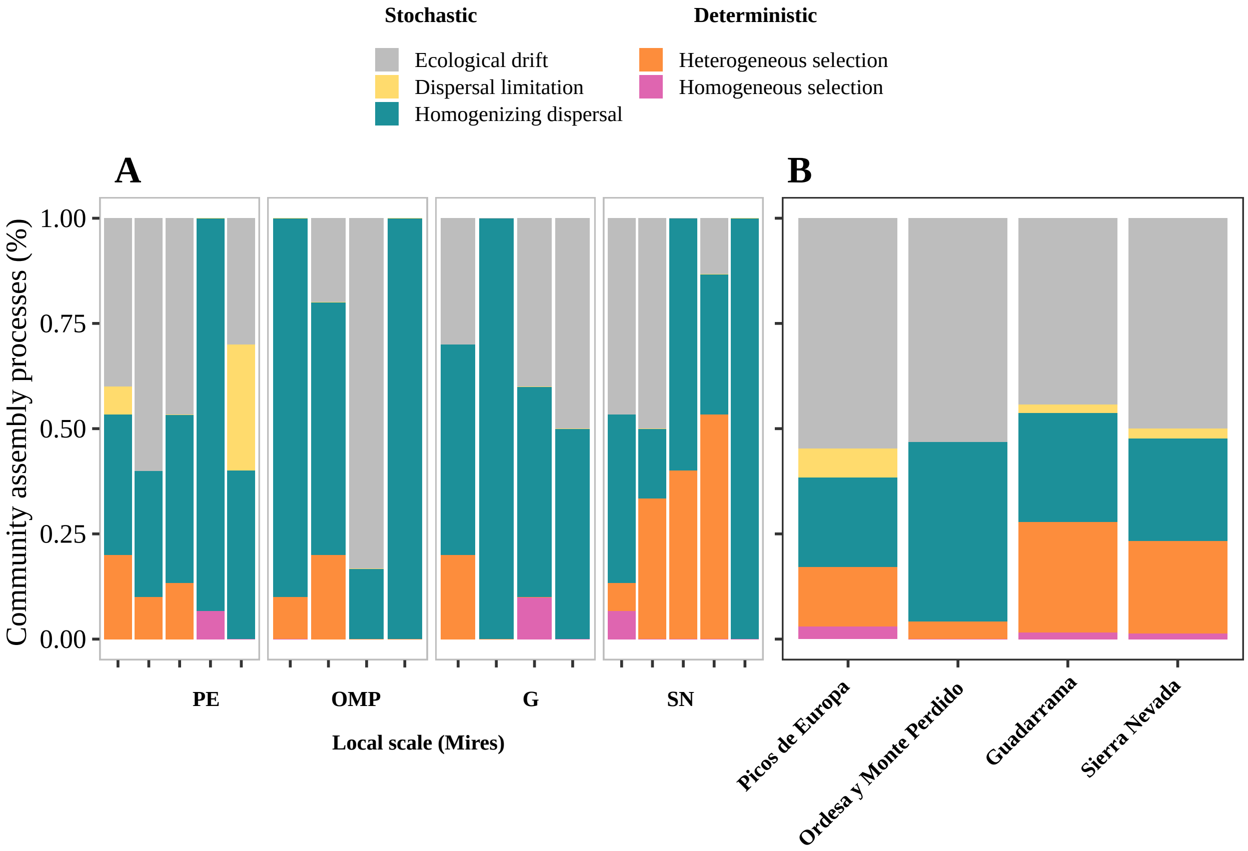
4 National parks inside analysis (Landscape scale)

4.1 Display panel Figure S3 B

```
cbp1 <- c("#bdbdbd","#FFDB6D" ,"#1c9099","#fd8d3c","#df65b0",
          "#56B4E9","#009E73","#293352","#F0E442","#ff0000")

agg.bRC_np$group<-factor(agg.bRC_np$Landscape.x,
                         c("Picos de Europa","Ordesa y Monte Perdido",
                           "Guadarrama", "Sierra Nevada"))
agg.bRC_np$proc<-factor(agg.bRC_np$process,
                        c("Ecological drift","Dispersal limitation",
                          "Homogenizing dispersal",
                          "Heterogeneous selection","Homogeneous selection"))

prop.bRC_np<-agg.bRC_np%>%
  ggplot(aes(x = group, y = RC.ij.bray, fill = proc)) +
  geom_col(position = "fill")+scale_fill_manual(values = c(cbp1[1:5]),
                                                name="Community assembly process")+
  labs(title="",x=NULL,y=expression(beta*"RC (%) "))+
  theme_bw()+
  theme(panel.grid.minor = element_blank(),panel.grid.major = element_blank(),
        axis.text.x = element_blank(),axis.title.x = element_blank(),
        axis.title.y = element_blank(),
        axis.text.y = element_blank(),  text = element_text(family="serif"))+
  theme(plot.margin = unit(c(0.1,0.15, 0.1,0), "cm"))+
  theme(legend.key.size = unit(1, "lines"),
        legend.margin=margin(0,1,0,1),legend.box.margin=margin(-10,-10,-10,-10),
        legend.position = "none", legend.justification="bottom",
        legend.box = "horizontal",
        legend.key = element_rect(color = NA, fill = NA),
        legend.spacing.x = unit(0.3,"cm"),legend.text=element_text(size=6))
```

4.2 Display Figure S2

Figure S2: A different approach, grouping multiple landscape estimates, shows the variability in the assembly processes of diatom communities at the regional scale. Representation of the various combinations of national parks (A, B, C). Percentage of community assembly processes in the six different combinations (D, E, F). The code represented the national parks' names: Ordesa y Monte Perdido (OMP), Guadarrama (G), Picos de Europa (PE), and Sierra Nevada (SN).



5 Comparison between pairs national parks

```
comm_p<-t(sqrt(Abun_rel[8:360]))
TaxonID_p<-row.names(comm_p)
meta.groupi_p<-data.frame(cluster=Abun_rel[,c("Landscape")])
row.names(meta.groupi_p)<-Abun_rel$Station

#taxonomic normalized stochasticity ratio
set.seed(1)
tnst_p=tNST(comm=t(comm_p),meta.com=NULL,group=meta.groupi_p,
            dist.method="bray", abundance.weighted=TRUE, rand=1000,
            output.rand=TRUE, nworker=8, LB=FALSE, null.model="PF",
            between.group=TRUE, SES=TRUE, RC=TRUE)

#to extract only the comparisons between each national park
nst.temp_p<-tnst_p[["index.pair"]]
nst.temp_p$index=paste0(nst.temp_p$name1,".",nst.temp_p$name2)

#RC percentage
nst.temp_p$category<-"bRC<|0.95|"
nst.temp_p$category[nst.temp_p$RC.bray>0.95]<-"bRC>0.95"
nst.temp_p$category[nst.temp_p$RC.bray< (-0.95)]<-"bRC<-0.95"

#To extract stochastic data
```

```
stoch_cate_p=nst.temp_p[,c(1,2,12,13,14)]
stoch_cate_p$proc<-"Ecological drift"
stoch_cate_p$proc[stoch_cate_p$RC.bray>0.95]<-"Dispersal limitation"
stoch_cate_p$proc[stoch_cate_p$RC.bray< (-0.95)]<-"Homogenizing dispersal"

xy_p <- t(combn(colnames(NTI.out$index), 2))
tmp_p = data.frame(xy_p, BNTI = NTI.out$index[xy_p])
tmp_merge_p=merge(tmp_p,metadata[,1:4],by.x="X1",by.y="Station")
indice_p=merge(tmp_merge_p,metadata[,1:4],by.x="X2",by.y="Station")
indice_p=indice_p[indice_p$Landscape.x!=indice_p$Landscape.y,]
indice_p$proc="Stochastic"
indice_p$proc[indice_p$BNTI< (-2)]="Homogeneous selection"
indice_p$proc[indice_p$BNTI>2]="Heterogeneous selection"
indice_p$index=paste0(indice_p$X2,".",indice_p$X1)

procesos_p=merge(indice_p,stoch_cate_p,by="index")

procesos_p$process<-NA
procesos_p$process=procesos_p$proc.x
procesos_p$process=ifelse(procesos_p$process=="Stochastic",
                          procesos_p$proc.y,procesos_p$proc.x)
procesos_p$index.np=paste0(procesos_p$Landscape.x,".",procesos_p$Landscape.y)
unique(procesos_p$index.np)
```

```
## [1] "Sierra Nevada.Guadarrama"
## [2] "Ordesa y Monte Perdido.Guadarrama"
## [3] "Picos de Europa.Guadarrama"
## [4] "Picos de Europa.Sierra Nevada"
## [5] "Sierra Nevada.Ordesa y Monte Perdido"
## [6] "Picos de Europa.Ordesa y Monte Perdido"
```

```
agg.bRC_p<-aggregate(RC.bray~process+index.np,data=procesos_p,length)
agg.bRC_p$category<-as.factor(agg.bRC_p$process) #Set it as factor

#percentage
gg.tmp_p<-aggregate(RC.bray~index.np,data=procesos_p,length)
gg.bRC_percentage_p<-merge(agg.bRC_p,gg.tmp_p,by="index.np")

gg.bRC_percentage_p$perc.RC<-
  gg.bRC_percentage_p$RC.bray.x/gg.bRC_percentage_p$RC.bray.y
```

5.1 Display panel Figure 3B

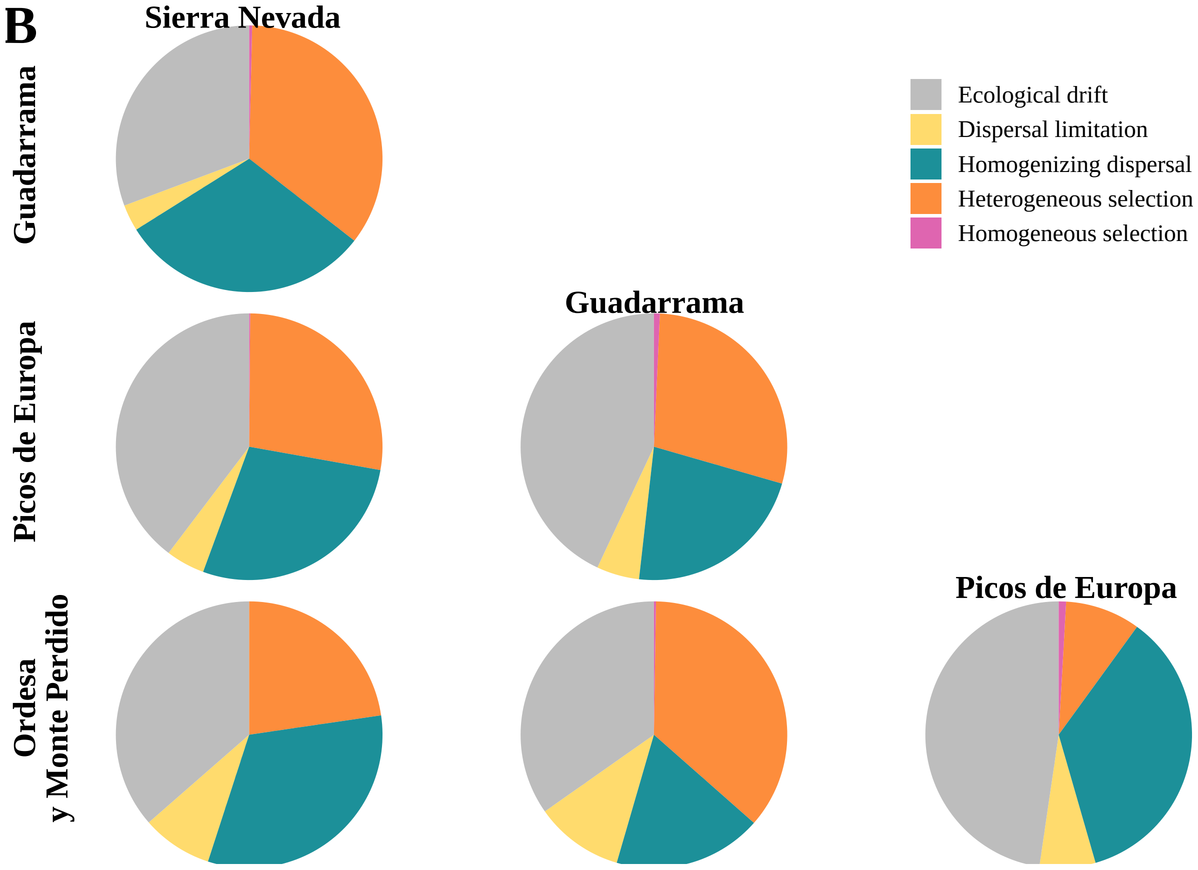
```
cbp1 <- c("#bdbdbd","#FFDB6D" ,"#1c9099","#fd8d3c","#df65b0",
          "#56B4E9","#009E73","#293352","#F0E442","#ff0000")

agg.bRC_p$group<-factor(agg.bRC_p$index.np,
                        c("Picos de Europa.Ordesa y Monte Perdido",
                          "Picos de Europa.Guadarrama",
                          "Picos de Europa.Sierra Nevada",
                          "Ordesa y Monte Perdido.Guadarrama",
                          "Sierra Nevada.Ordesa y Monte Perdido",
                          "Sierra Nevada.Guadarrama"))
```

```r
agg.bRC_p$proc<-factor(agg.bRC_p$process,
                       c("Ecological drift","Dispersal limitation",
                         "Homogenizing dispersal",
                         "Heterogeneous selection","Homogeneous selection"))


RC_pe_omp=agg.bRC_p[agg.bRC_p$group=="Picos de Europa.Ordesa y Monte Perdido",]%>%
  ggplot(aes(x="" ,y = RC.bray, fill = proc)) +
  geom_bar(stat="identity", width=1) +
  coord_polar("y", start=0)+
  scale_fill_manual(values = c(cbp1[1:5]),name="")+
  theme_minimal()+ theme_void()+
  theme(legend.position = "none")+theme(plot.margin = unit(c(-0.1,-.3,-.5,-.1),"cm"))



RC_pe_g=agg.bRC_p[agg.bRC_p$group=="Picos de Europa.Guadarrama",]%>%
  ggplot(aes(x="" ,y = RC.bray, fill = proc)) +
  geom_bar(stat="identity", width=1) +
  coord_polar("y", start=0)+
  scale_fill_manual(values = c(cbp1[1:5]),name="")+
  theme_minimal()+ theme_void()+
  theme(legend.position = "none")+theme(plot.margin = unit(c(-0.1,-.3,-.5,-.1),"cm"))



RC_pe_sn=agg.bRC_p[agg.bRC_p$group=="Picos de Europa.Sierra Nevada",]%>%
  ggplot(aes(x="" ,y = RC.bray, fill = proc)) +
  geom_bar(stat="identity", width=1) +
  coord_polar("y", start=0)+
  scale_fill_manual(values = c(cbp1[1:5]),name="")+
  theme_minimal()+ theme_void()+
  theme(legend.position = "none")+theme(plot.margin = unit(c(-0.1,-.3,-.5,-.1),"cm"))



RC_omp_g=agg.bRC_p[agg.bRC_p$group=="Ordesa y Monte Perdido.Guadarrama",]%>%
  ggplot(aes(x="" ,y = RC.bray, fill = proc)) +
  geom_bar(stat="identity", width=1) +
  coord_polar("y", start=0)+
  scale_fill_manual(values = c(cbp1[1:5]),name="")+
  theme_minimal()+ theme_void()+
  theme(legend.position = "none")+theme(plot.margin = unit(c(-0.1,-.3,-.5,-.1),"cm"))



RC_sn_omp=agg.bRC_p[agg.bRC_p$group=="Sierra Nevada.Ordesa y Monte Perdido",]%>%
  ggplot(aes(x="" ,y = RC.bray, fill = proc)) +
  geom_bar(stat="identity", width=1) +
  coord_polar("y", start=0)+
  scale_fill_manual(values = c(cbp1[1:5]),name="")+
  theme_minimal()+ theme_void()+
  theme(legend.position = "none")+theme(plot.margin = unit(c(-0.1,-.3,-.5,-.1),"cm"))
```

```
RC_sn_g=agg.bRC_p[agg.bRC_p$group=="Sierra Nevada.Guadarrama",]%>%
  ggplot(aes(x="" ,y = RC.bray, fill = proc)) +
  geom_bar(stat="identity", width=1) +
  coord_polar("y", start=0)+
  scale_fill_manual(values = c(cbp1[1:5]),name="")+
  theme_minimal()+ theme_void()+
  theme(text = element_text(family="serif"))+
  theme(legend.position = "right",legend.key.size = unit(0.9, "lines"),
        legend.margin=margin(0,1,0,1),legend.box.margin=margin(-10,-10,-10,-10),
        legend.box = "horizontal",legend.key = element_rect(color = NA, fill = NA),
        legend.spacing.x = unit(0.1,"cm"),legend.text=element_text(size=9),
        plot.margin = unit(c(-0.1,-.3,-.5,-.1),"cm"))
legenda_sn_g=get_legend(RC_sn_g)
RC_sn_g=RC_sn_g+theme(legend.position = "none")
```

5.2 Figure 3

Figure 3: Pairwise comparison in assembly processes between all landscapes (national parks). A) Description of the pairwise comparison conducted between national parks. B) The percentage of assembly processes between national parks.



6 Six different combination of national parks to calculate regional community assembly

6.1 First combination

```
comm_pp<-t(sqrt(Abun_rel[8:360]))
TaxonID_pp<-row.names(comm_pp)
meta.groupi_pp<-data.frame(cluster=Abun_rel[,c("Regional_1")])
row.names(meta.groupi_pp)<-Abun_rel$Station

#taxonomic normalized stochasticity ratio
set.seed(1)
tnst_pp=tNST(comm=t(comm_pp),meta.com=NULL,group=meta.groupi_pp,
             dist.method="bray", abundance.weighted=TRUE, rand=1000,
             output.rand=TRUE, nworker=8, LB=FALSE, null.model="PF",
             between.group=TRUE, SES=TRUE, RC=TRUE)

#to extract each combination of national parks
nst.temp_pp<-tnst_pp[["index.pair.grp"]]
nst.temp_pp$index=paste0(nst.temp_pp$name1,".",nst.temp_pp$name2)

#RC percentage
nst.temp_pp$category<-"bRC<|0.95|"
nst.temp_pp$category[nst.temp_pp$RC.ij.bray>0.95]<-"bRC>0.95"
nst.temp_pp$category[nst.temp_pp$RC.ij.bray< (-0.95)]<-"bRC<-0.95"

#To extract stochastic data
stoch_cate_pp=nst.temp_pp[,c(2,3,10,11,12)]
stoch_cate_pp$proc<-"Ecological drift"
stoch_cate_pp$proc[stoch_cate_pp$RC.ij.bray>0.95]<-"Dispersal limitation"
stoch_cate_pp$proc[stoch_cate_pp$RC.ij.bray< (-0.95)]<-"Homogenizing dispersal"

xy_pp <- t(combn(colnames(NTI.out$index), 2))
tmp_pp = data.frame(xy_pp, BNTI = NTI.out$index[xy_pp])
tmp_merge_pp=merge(tmp_pp,Abun_rel[,4:7],by.x="X1",by.y="Station")
indice_pp=merge(tmp_merge_pp,Abun_rel[,4:7],by.x="X2",by.y="Station")
indice_pp=indice_pp[indice_pp$Regional_1.x==indice_pp$Regional_1.y,]
indice_pp$proc="Stochastic"
indice_pp$proc[indice_pp$BNTI< (-2)]="Homogeneous selection"
indice_pp$proc[indice_pp$BNTI>2]="Heterogeneous selection"
indice_pp$index=paste0(indice_pp$X2,".",indice_pp$X1)

procesos_pp=merge(indice_pp,stoch_cate_pp,by="index")

procesos_pp$process<-NA
procesos_pp$process=procesos_pp$proc.x
procesos_pp$process=ifelse(procesos_pp$process=="Stochastic",
                           procesos_pp$proc.y,procesos_pp$proc.x)
procesos_pp$index.np=paste0(procesos_pp$Regional_1.x,".",
                            procesos_pp$Regional_1.y)
unique(procesos_pp$index.np)
```

```
## [1] "OMP_G.OMP_G" "PE_SN.PE_SN"
```

```
agg.bRC_pp<-aggregate(RC.ij.bray~process+index.np,data=procesos_pp,length)
agg.bRC_pp$category<-as.factor(agg.bRC_pp$process) #Set it as factor

#percentage
gg.tmp_pp<-aggregate(RC.ij.bray~index.np,data=procesos_pp,length)
gg.bRC_percentage_pp<-merge(agg.bRC_pp,gg.tmp_pp,by="index.np")

gg.bRC_percentage_pp$perc.RC<-
  gg.bRC_percentage_pp$RC.ij.bray.x/gg.bRC_percentage_pp$RC.ij.bray.y
```

6.1.1 Display panel Figure S2 D

```
cbp1 <- c("#bdbdbd","#FFDB6D" ,"#1c9099","#fd8d3c","#df65b0",
          "#56B4E9","#009E73","#293352","#F0E442","#ff0000")

agg.bRC_pp$group<-factor(agg.bRC_pp$index.np,
                         c("OMP_G.OMP_G","PE_SN.PE_SN"))
agg.bRC_pp$proc<-factor(agg.bRC_pp$process,
                        c("Ecological drift","Dispersal limitation",
                          "Homogenizing dispersal",
                          "Heterogeneous selection","Homogeneous selection"))

prop.bRC_pp<-agg.bRC_pp%>%
  ggplot(aes(x = group, y = RC.ij.bray, fill = proc)) +
  geom_col(position = "fill")+scale_fill_manual(values = c(cbp1[1:5]),name="")+
  labs(title="",x=NULL,y="Community assembly processes (%) ")+
  theme_bw()+
  theme(panel.grid.minor = element_blank(),panel.grid.major = element_blank(),
        axis.text.y = element_text(size=10, color="black",hjust=1,angle = 0),
        axis.text.x = element_blank(),axis.title.x = element_blank())+
  theme(text = element_text(family = "serif"))+
  theme(plot.margin = unit(c(0.1,0.15, 0.1,0), "cm"))+
  theme(legend.key.size = unit(1, "lines"),
        legend.margin=margin(0,1,0,1),legend.box.margin=margin(-10,-10,-10,-10),
        legend.position = "none",legend.box = "horizontal",
        legend.key = element_rect(color = NA, fill = NA),
        legend.spacing.x = unit(0.3,"cm"),legend.text=element_text(size=6))
```

6.2 Second combination

```
comm_pp2<-t(sqrt(Abun_rel[8:360]))
TaxonID_pp2<-row.names(comm_pp2)
meta.groupi_pp2<-data.frame(cluster=Abun_rel[,c("Regional_2")])
row.names(meta.groupi_pp2)<-Abun_rel$Station

#taxonomic normalized stochasticity ratio
set.seed(1)
tnst_pp2=tNST(comm=t(comm_pp2),meta.com=NULL,group=meta.groupi_pp2,
              dist.method="bray", abundance.weighted=TRUE, rand=1000,
              output.rand=TRUE, nworker=8, LB=FALSE, null.model="PF",
              between.group=TRUE, SES=TRUE, RC=TRUE)
```

```r
#to extract each combination of national parks
nst.temp_pp2<-tnst_pp2[["index.pair.grp"]]
nst.temp_pp2$index=paste0(nst.temp_pp2$name1,".",nst.temp_pp2$name2)

#RC percentage
nst.temp_pp2$category<-"bRC<|0.95|"
nst.temp_pp2$category[nst.temp_pp2$RC.ij.bray>0.95]<-"bRC>0.95"
nst.temp_pp2$category[nst.temp_pp2$RC.ij.bray< (-0.95)]<-"bRC<-0.95"

#To extract stochastic data
stoch_cate_pp2=nst.temp_pp2[,c(1,2,3,10,11,12)]
stoch_cate_pp2$proc<-"Ecological drift"
stoch_cate_pp2$proc[stoch_cate_pp2$RC.ij.bray>0.95]<-"Dispersal limitation"
stoch_cate_pp2$proc[stoch_cate_pp2$RC.ij.bray< (-0.95)]<-"Homogenizing dispersal"

xy_pp2 <- t(combn(colnames(NTI.out$index), 2))
tmp_pp2 = data.frame(xy_pp2, BNTI = NTI.out$index[xy_pp2])
tmp_merge_pp2=merge(tmp_pp2,Abun_rel[,3:7],by.x="X1",by.y="Station")
indice_pp2=merge(tmp_merge_pp2,Abun_rel[,3:7],by.x="X2",by.y="Station")
indice_pp2=indice_pp2[indice_pp2$Regional_2.x==indice_pp2$Regional_2.y,]
indice_pp2$proc="Stochastic"
indice_pp2$proc[indice_pp2$BNTI< (-2)]="Homogeneous selection"
indice_pp2$proc[indice_pp2$BNTI>2]="Heterogeneous selection"
indice_pp2$index=paste0(indice_pp2$X2,".",indice_pp2$X1)

procesos_pp2=merge(indice_pp2,stoch_cate_pp2,by="index")

procesos_pp2$process<-NA
procesos_pp2$process=procesos_pp2$proc.x
procesos_pp2$process=ifelse(procesos_pp2$process=="Stochastic",
                            procesos_pp2$proc.y,procesos_pp2$proc.x)
procesos_pp2$index.np=paste0(procesos_pp2$Regional_2.x,".",
                             procesos_pp2$Regional_2.y)
unique(procesos_pp2$index.np)
```

```
## [1] "SN_G.SN_G"     "PE_OMP.PE_OMP"
```

```r
agg.bRC_pp2<-aggregate(RC.ij.bray~process+index.np,data=procesos_pp2,length)
agg.bRC_pp2$category<-as.factor(agg.bRC_pp2$process) #Set it as factor

#percentage
gg.tmp_pp2<-aggregate(RC.ij.bray~index.np,data=procesos_pp2,length)
gg.bRC_percentage_pp2<-merge(agg.bRC_pp2,gg.tmp_pp2,by="index.np")

gg.bRC_percentage_pp2$perc.RC<-
  gg.bRC_percentage_pp2$RC.ij.bray.x/gg.bRC_percentage_pp2$RC.ij.bray.y
```

6.2.1 Display panel Figure S2 E

```r
cbp1 <- c("#bdbdbd","#FFDB6D" ,"#1c9099","#fd8d3c","#df65b0",
          "#56B4E9","#009E73","#293352","#F0E442","#ff0000")

agg.bRC_pp2$group<-factor(agg.bRC_pp2$index.np,
                          c("SN_G.SN_G","PE_OMP.PE_OMP"))
```

```r
agg.bRC_pp2$proc<-factor(agg.bRC_pp2$process,
                         c("Ecological drift","Dispersal limitation",
                           "Homogenizing dispersal",
                           "Heterogeneous selection","Homogeneous selection"))

prop.bRC_pp2<-agg.bRC_pp2%>%
  ggplot(aes(x = group, y = RC.ij.bray, fill = proc)) +
  geom_col(position = "fill")+scale_fill_manual(values = c(cbp1[1:5]),name="")+
  labs(title="",x=NULL,y="Percentage (%) ")+
  theme_bw()+
  theme(panel.grid.minor = element_blank(),panel.grid.major = element_blank(),
        axis.text.y = element_text(size=10, color="black",hjust=1,angle = 0),
        axis.text.x = element_blank(),axis.title.x = element_blank(),
        axis.title.y = element_blank())+
  theme(text = element_text(family = "serif"))+
  theme(plot.margin = unit(c(0.1,0.15, 0.1,0), "cm"))+
  theme(legend.key.size = unit(1, "lines"),
        legend.margin=margin(0,1,0,1),legend.box.margin=margin(-10,-10,-10,-10),
        legend.position = "none",legend.box = "horizontal",
        legend.key = element_rect(color = NA, fill = NA),
        legend.spacing.x = unit(0.3,"cm"),legend.text=element_text(size=6))
```

6.3 Third combination

```r
comm_pp3<-t(sqrt(Abun_rel[8:360]))
TaxonID_pp3<-row.names(comm_pp3)
meta.groupi_pp3<-data.frame(cluster=Abun_rel[,c("Regional_3")])
row.names(meta.groupi_pp3)<-Abun_rel$Station

#taxonomic normalized stochasticity ratio
set.seed(1)
tnst_pp3=tNST(comm=t(comm_pp3),meta.com=NULL,group=meta.groupi_pp3,
              dist.method="bray", abundance.weighted=TRUE, rand=1000,
              output.rand=TRUE, nworker=8, LB=FALSE, null.model="PF",
              between.group=TRUE, SES=TRUE, RC=TRUE)

nst.temp_pp3<-tnst_pp3[["index.pair.grp"]]#extraemos dentro de los grupo
nst.temp_pp3$index=paste0(nst.temp_pp3$name1,".",nst.temp_pp3$name2)

#RC percentage
nst.temp_pp3$category<-"bRC<|0.95|"
nst.temp_pp3$category[nst.temp_pp3$RC.ij.bray>0.95]<-"bRC>0.95"
nst.temp_pp3$category[nst.temp_pp3$RC.ij.bray< (-0.95)]<-"bRC<-0.95"

#To extract stochastic data
stoch_cate_pp3=nst.temp_pp3[,c(1,2,3,10,11,12)]
stoch_cate_pp3$proc<-"Ecological drift"
stoch_cate_pp3$proc[stoch_cate_pp3$RC.ij.bray>0.95]<-"Dispersal limitation"
stoch_cate_pp3$proc[stoch_cate_pp3$RC.ij.bray< (-0.95)]<-"Homogenizing dispersal"

xy_pp3 <- t(combn(colnames(NTI.out$index), 2))
tmp_pp3 = data.frame(xy_pp3, BNTI = NTI.out$index[xy_pp3])
tmp_merge_pp3=merge(tmp_pp3,Abun_rel[,2:7],by.x="X1",by.y="Station")
```

```
indice_pp3=merge(tmp_merge_pp3,Abun_rel[,2:7],by.x="X2",by.y="Station")
indice_pp3=indice_pp3[indice_pp3$Regional_3.x==indice_pp3$Regional_3.y,]
indice_pp3$proc="Stochastic"
indice_pp3$proc[indice_pp3$BNTI< (-2)]="Homogeneous selection"
indice_pp3$proc[indice_pp3$BNTI>2]="Heterogeneous selection"
indice_pp3$index=paste0(indice_pp3$X2,".",indice_pp3$X1)

procesos_pp3=merge(indice_pp3,stoch_cate_pp3,by="index")

procesos_pp3$process<-NA
procesos_pp3$process=procesos_pp3$proc.x
procesos_pp3$process=ifelse(procesos_pp3$process=="Stochastic",
                            procesos_pp3$proc.y,procesos_pp3$proc.x)
procesos_pp3$index.np=paste0(procesos_pp3$Regional_3.x,".",
                             procesos_pp3$Regional_3.y)
unique(procesos_pp3$index.np)
```

```
## [1] "PE_G.PE_G"     "SN_OMP.SN_OMP"
```

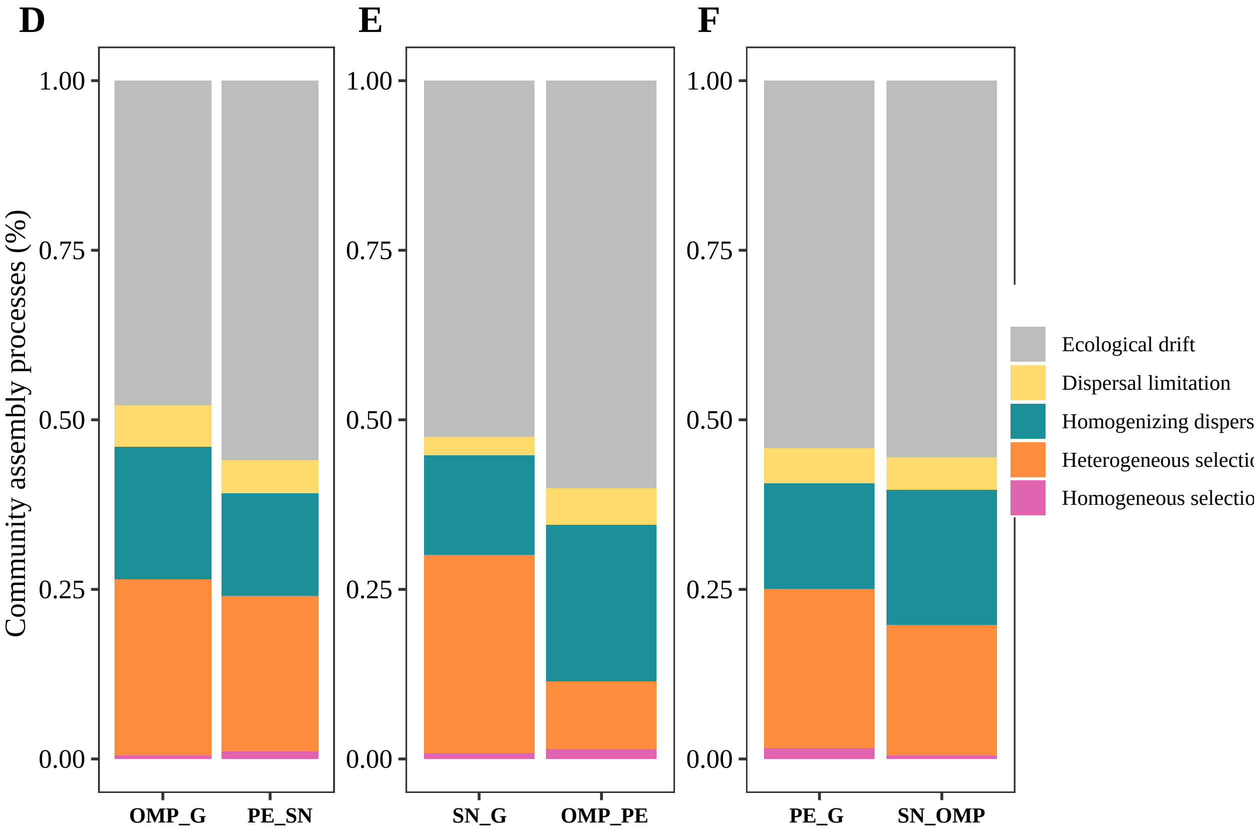
```
agg.bRC_pp3<-aggregate(RC.ij.bray~process+index.np,data=procesos_pp3,length)
agg.bRC_pp3$category<-as.factor(agg.bRC_pp3$process) #Set it as factor

#percentage
gg.tmp_pp3<-aggregate(RC.ij.bray~index.np,data=procesos_pp3,length)
gg.bRC_percentage_pp3<-merge(agg.bRC_pp3,gg.tmp_pp3,by="index.np")

gg.bRC_percentage_pp3$perc.RC<-
  gg.bRC_percentage_pp3$RC.ij.bray.x/gg.bRC_percentage_pp3$RC.ij.bray.y
```

6.3.1 Display panel Figure S2 F

```
cbp1 <- c("#bdbdbd","#FFDB6D" ,"#1c9099","#fd8d3c","#df65b0",
          "#56B4E9","#009E73","#293352","#F0E442","#ff0000")

agg.bRC_pp3$group<-factor(agg.bRC_pp3$index.np,
                          c("PE_G.PE_G","SN_OMP.SN_OMP"))
agg.bRC_pp3$proc<-factor(agg.bRC_pp3$process,
                         c("Ecological drift","Dispersal limitation",
                           "Homogenizing dispersal",
                           "Heterogeneous selection","Homogeneous selection"))

prop.bRC_pp3<-agg.bRC_pp3%>%
  ggplot(aes(x = group, y = RC.ij.bray, fill = proc)) +
  geom_col(position = "fill")+scale_fill_manual(values = c(cbp1[1:5]),name="")+
  labs(title="",x=NULL,y="Percentage (%) ")+
  theme_bw()+
  theme(panel.grid.minor = element_blank(),panel.grid.major = element_blank(),
        axis.text.y = element_text(size=10, color="black",hjust=1,angle = 0),
        axis.text.x = element_blank(),axis.title.x = element_blank(),
        axis.title.y = element_blank())+
  theme(text = element_text(family = "serif"))+
  theme(plot.margin = unit(c(0.1,0.15, 0.1,0), "cm"))+
  theme(legend.key.size = unit(1, "lines"),
        legend.margin=margin(0,1,0,1),legend.box.margin=margin(-10,-10,-10,-10),
        legend.position = "right",legend.box = "horizontal",
        legend.key = element_rect(color = NA, fill = NA),
        legend.spacing.x = unit(0.3,"cm"),legend.text=element_text(size=8))
legenda_prop.bRC_pp3=get_legend(prop.bRC_pp3)
prop.bRC_pp3=prop.bRC_pp3+theme(legend.position = "none")
```

6.4 Display Figure S2

Figure S2: A different approach, grouping multiple landscape estimates, shows the variability in the assembly processes of diatom communities at the regional scale. Representation of the various combinations of national parks (A, B, C). Percentage of community assembly processes in the six different combinations (D, E, F). The code represented the national parks' names: Ordesa y Monte Perdido (OMP), Guadarrama (G), Picos de Europa (PE), and Sierra Nevada (SN).



7 Percentages of community assembly processes

```
percentage=data.frame(read_xlsx("data\\Assembly.xlsx",sheet="Percentage",na=""))

#Mean and standard deviation
tapply(percentage$perc.RC,interaction(percentage$category),mean)
```

```
##    Dispersal limitation        Ecological drift Heterogeneous selection
##              0.02129647              0.36171625              0.14656093
##   Homogeneous selection  Homogenizing dispersal
##              0.02293006              0.44749629
```

```r
tapply(percentage$perc.RC,interaction(percentage$category),sd)
```

```
##    Dispersal limitation        Ecological drift Heterogeneous selection 
##              0.04122051              0.22029302              0.12664971 
##   Homogeneous selection  Homogenizing dispersal 
##              0.03159312              0.28510311
```

```r
tapply(percentage$perc.RC,interaction(percentage$proc),mean)
```

```
## Deterministic    Stochastic 
##     0.1055024     0.2688316
```

```r
tapply(percentage$perc.RC,interaction(percentage$proc),sd)
```

```
## Deterministic    Stochastic 
##     0.1407632     0.2795182
```

```r
tapply(percentage$perc.RC,interaction(percentage$group,percentage$proc),mean)
```

```
## Landscape.Deterministic     Local.Deterministic  Regional.Deterministic 
##              0.20692894              0.06944444              0.12070174 
##    Landscape.Stochastic        Local.Stochastic     Regional.Stochastic 
##              0.19153130              0.28703704              0.25286551
```

```r
tapply(percentage$perc.RC,interaction(percentage$group,percentage$proc),sd)
```

```
## Landscape.Deterministic     Local.Deterministic  Regional.Deterministic 
##               0.2031950               0.1093632               0.1187036 
##    Landscape.Stochastic        Local.Stochastic     Regional.Stochastic 
##               0.1837350               0.3108396               0.2146375
```

```r
tapply(percentage$perc.RC,interaction(percentage$group,percentage$category),mean)
```

```
##    Landscape.Dispersal limitation        Local.Dispersal limitation 
##                        0.01982735                        0.01481481 
##     Regional.Dispersal limitation        Landscape.Ecological drift 
##                        0.04172087                        0.48156502 
##            Local.Ecological drift         Regional.Ecological drift 
##                        0.27777778                        0.53363247 
## Landscape.Heterogeneous selection     Local.Heterogeneous selection 
##                        0.17753411                        0.11296296 
##  Regional.Heterogeneous selection   Landscape.Homogeneous selection 
##                        0.22670604                        0.02179755 
##       Local.Homogeneous selection    Regional.Homogeneous selection 
##                        0.02592593                        0.01469745 
##  Landscape.Homogenizing dispersal      Local.Homogenizing dispersal 
##                        0.29927597                        0.56851852 
##   Regional.Homogenizing dispersal 
##                        0.18324318
```

```r
tapply(percentage$perc.RC,interaction(percentage$group,percentage$category),sd)
```

```
##    Landscape.Dispersal limitation        Local.Dispersal limitation 
##                       0.021662327                       0.048805633 
##     Regional.Dispersal limitation        Landscape.Ecological drift 
##                       0.011892392                       0.027467275 
##            Local.Ecological drift         Regional.Ecological drift 
##                       0.235147013                       0.035810619
## Landscape.Heterogeneous selection     Local.Heterogeneous selection
##                       0.092398056                       0.138214047
##  Regional.Heterogeneous selection   Landscape.Homogeneous selection
##                       0.063282165                       0.013748501
##       Local.Homogeneous selection    Regional.Homogeneous selection
##                       0.038865539                       0.005098309
##  Landscape.Homogenizing dispersal      Local.Homogenizing dispersal
##                       0.113788250                       0.284793750
##   Regional.Homogenizing dispersal
##                       0.036059804
```

```
#Normality test and variance homogenization of variance
tapply(sqrt(percentage$perc.RC),interaction(percentage$group,percentage$proc),shapiro.test)
```

```
## $Landscape.Deterministic
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.87965, p-value = 0.103
##
##
## $Local.Deterministic
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.77991, p-value = 6.658e-06
##
##
## $Regional.Deterministic
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.80857, p-value = 0.01175
##
##
## $Landscape.Stochastic
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.91358, p-value = 0.3418
##
##
## $Local.Stochastic
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.84751, p-value = 6.778e-06
##
##
## $Regional.Stochastic
```

```
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.87466, p-value = 0.02125
```

```
bartlett.test(sqrt(percentage$perc.RC)~interaction(percentage$group,percentage$proc))
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  sqrt(percentage$perc.RC) by interaction(percentage$group, percentage$proc)
## Bartlett's K-squared = 17.307, df = 5, p-value = 0.003953
```

```
#make subgroups to evaluate statistical differences between
#stochastic and deterministic processes
mire=percentage[1:90,]
n_park=percentage[91:110,]
p_iberica=percentage[11:140,]

#mire
tapply(mire$perc.RC,interaction(mire$proc),shapiro.test)
```

```
## $Deterministic
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.66709, p-value = 8.936e-08
##
##
## $Stochastic
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.8442, p-value = 5.49e-06
```

```
bartlett.test(mire$perc.RC,mire$proc)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  mire$perc.RC and mire$proc
## Bartlett's K-squared = 34.996, df = 1, p-value = 3.304e-09
```

```
wilcox.test(mire$perc.RC~mire$proc)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  mire$perc.RC by mire$proc
## W = 627, p-value = 0.002942
## alternative hypothesis: true location shift is not equal to 0
```

```
#Landscape
tapply(n_park$perc.RC,interaction(n_park$proc),shapiro.test)
```

```
## $Deterministic
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.82839, p-value = 0.02224
##
##
## $Stochastic
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.87215, p-value = 0.1296
```

```
bartlett.test(n_park$perc.RC,n_park$proc)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  n_park$perc.RC and n_park$proc
## Bartlett's K-squared = 0.084499, df = 1, p-value = 0.7713
```

```
t.test(n_park$perc.RC~n_park$proc)
```

```
##
##  Welch Two Sample t-test
##
## data:  n_park$perc.RC by n_park$proc
## t = 0.17774, df = 17.779, p-value = 0.8609
## alternative hypothesis: true difference in means between group Deterministic and group Stochastic is
## 95 percent confidence interval:
##  -0.1667634  0.1975587
## sample estimates:
## mean in group Deterministic    mean in group Stochastic
##                   0.2069289                   0.1915313
```

```
#Regional
tapply(p_iberica$perc.RC,interaction(p_iberica$proc),shapiro.test)
```

```
## $Deterministic
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.7768, p-value = 9.763e-08
##
##
## $Stochastic
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.87032, p-value = 1.601e-06
```

```
bartlett.test(p_iberica$perc.RC,p_iberica$proc)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  p_iberica$perc.RC and p_iberica$proc
## Bartlett's K-squared = 21.839, df = 1, p-value = 2.965e-06
```

```
wilcox.test(p_iberica$perc.RC~p_iberica$proc)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  p_iberica$perc.RC by p_iberica$proc
## W = 1456.5, p-value = 0.003946
## alternative hypothesis: true location shift is not equal to 0
```

```
#inside scale evaluate the different processes
#mire
k<-kruskal.test(mire$perc.RC~mire$category)
k
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  mire$perc.RC by mire$category
## Kruskal-Wallis chi-squared = 51.199, df = 4, p-value = 2.028e-10
```

```
kruskalmc(mire$perc.RC~mire$category)
```

```
## Multiple comparison test after Kruskal-Wallis
## alpha: 0.05
## Comparisons
##                                                  obs.dif critical.dif
## Dispersal limitation-Ecological drift          31.666667     24.44431
## Dispersal limitation-Heterogeneous selection   18.416667     24.44431
## Dispersal limitation-Homogeneous selection      5.472222     24.44431
## Dispersal limitation-Homogenizing dispersal    52.083333     24.44431
## Ecological drift-Heterogeneous selection       13.250000     24.44431
## Ecological drift-Homogeneous selection         26.194444     24.44431
## Ecological drift-Homogenizing dispersal        20.416667     24.44431
## Heterogeneous selection-Homogeneous selection  12.944444     24.44431
## Heterogeneous selection-Homogenizing dispersal 33.666667     24.44431
## Homogeneous selection-Homogenizing dispersal   46.611111     24.44431
##                                                stat.signif
## Dispersal limitation-Ecological drift                 TRUE
## Dispersal limitation-Heterogeneous selection         FALSE
## Dispersal limitation-Homogeneous selection           FALSE
## Dispersal limitation-Homogenizing dispersal           TRUE
## Ecological drift-Heterogeneous selection             FALSE
## Ecological drift-Homogeneous selection                TRUE
## Ecological drift-Homogenizing dispersal              FALSE
## Heterogeneous selection-Homogeneous selection        FALSE
## Heterogeneous selection-Homogenizing dispersal        TRUE
## Homogeneous selection-Homogenizing dispersal          TRUE
```

```r
#Landscape
k<-kruskal.test(n_park$perc.RC~n_park$category)
k
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  n_park$perc.RC by n_park$category
## Kruskal-Wallis chi-squared = 16.371, df = 4, p-value = 0.002559
```

```r
kruskalmc(n_park$perc.RC~n_park$category)
```

```
## Multiple comparison test after Kruskal-Wallis
## alpha: 0.05
## Comparisons
##                                                obs.dif critical.dif stat.signif
## Dispersal limitation-Ecological drift            13.75     11.74266        TRUE
## Dispersal limitation-Heterogeneous selection      6.25     11.74266       FALSE
## Dispersal limitation-Homogeneous selection        0.25     11.74266       FALSE
## Dispersal limitation-Homogenizing dispersal       9.75     11.74266       FALSE
## Ecological drift-Heterogeneous selection          7.50     11.74266       FALSE
## Ecological drift-Homogeneous selection           13.50     11.74266        TRUE
## Ecological drift-Homogenizing dispersal           4.00     11.74266       FALSE
## Heterogeneous selection-Homogeneous selection     6.00     11.74266       FALSE
## Heterogeneous selection-Homogenizing dispersal    3.50     11.74266       FALSE
## Homogeneous selection-Homogenizing dispersal      9.50     11.74266       FALSE
```

```r
#Regional
k<-kruskal.test(p_iberica$perc.RC~p_iberica$category)
k
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  p_iberica$perc.RC by p_iberica$category
## Kruskal-Wallis chi-squared = 70.665, df = 4, p-value = 1.643e-14
```

```r
kruskalmc(p_iberica$perc.RC~p_iberica$category)
```

```
## Multiple comparison test after Kruskal-Wallis
## alpha: 0.05
## Comparisons
##                                                  obs.dif critical.dif
## Dispersal limitation-Ecological drift          60.653846     29.32869
## Dispersal limitation-Heterogeneous selection   32.153846     29.32869
## Dispersal limitation-Homogeneous selection      4.750000     29.32869
## Dispersal limitation-Homogenizing dispersal    66.576923     29.32869
## Ecological drift-Heterogeneous selection       28.500000     29.32869
## Ecological drift-Homogeneous selection         55.903846     29.32869
## Ecological drift-Homogenizing dispersal         5.923077     29.32869
## Heterogeneous selection-Homogeneous selection  27.403846     29.32869
## Heterogeneous selection-Homogenizing dispersal 34.423077     29.32869
## Homogeneous selection-Homogenizing dispersal   61.826923     29.32869
##                                                stat.signif
## Dispersal limitation-Ecological drift                 TRUE
## Dispersal limitation-Heterogeneous selection          TRUE
## Dispersal limitation-Homogeneous selection           FALSE
## Dispersal limitation-Homogenizing dispersal           TRUE
## Ecological drift-Heterogeneous selection             FALSE
## Ecological drift-Homogeneous selection                TRUE
## Ecological drift-Homogenizing dispersal              FALSE
## Heterogeneous selection-Homogeneous selection        FALSE
## Heterogeneous selection-Homogenizing dispersal        TRUE
## Homogeneous selection-Homogenizing dispersal          TRUE
```

```
percentage$category<-factor(percentage$category,
                            c("Ecological drift","Dispersal limitation",
                              "Homogenizing dispersal",
                              "Heterogeneous selection",
                              "Homogeneous selection"))

percentage$group<-factor(percentage$group,
                         c("Local","Landscape","Regional"))
```

7.1 Display panel Figure 2B

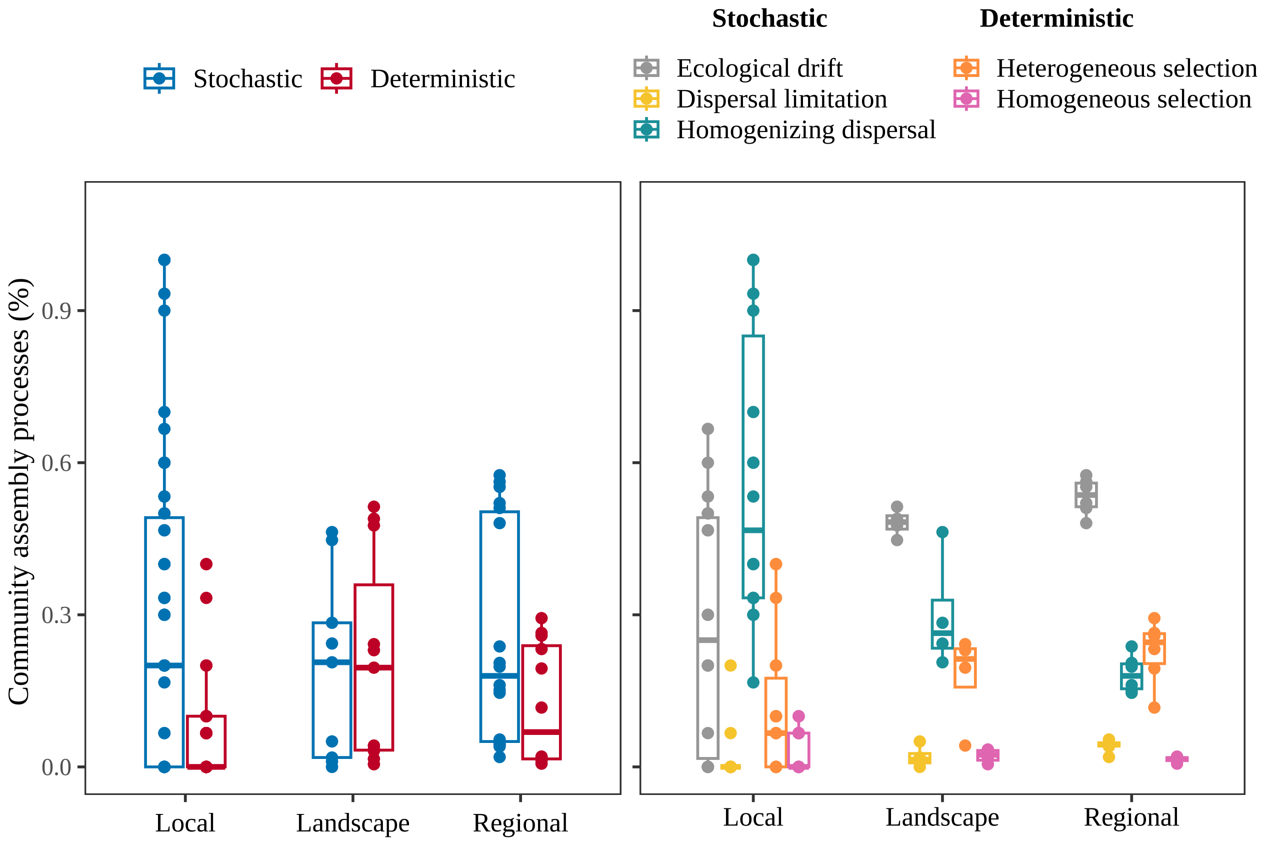
```
cbp1 <- c("#969696","#f5c32c" ,"#1c9099","#fd8d3c","#df65b0",
          "#56B4E9","#009E73","#293352","#F0E442","#ff0000")

perc_tur=ggplot(percentage,aes(x=group, y = perc.RC,colour=category)) +
  geom_boxplot(width=0.6,outlier.colour =NA)+
  geom_jitter(position=position_dodge(0.6))+
  scale_color_manual(values = c(cbp1[c(1:5)]),name="",
                     labels=c(expression("Ecological drift"),
                              expression("Dispersal limitation"),
                              expression("Homogenizing dispersal"),
                              expression("Heterogeneous selection"),
                              expression("Homogeneous selection")))+
  labs(title="",y="Percentage (%) ",x="Assembly process")+ylim(0,1.1)+
  theme_bw()+
  theme(panel.grid.minor = element_blank(),panel.grid.major = element_blank(),
        axis.text.x = element_text(size=10, color="black",angle =0,hjust=0.5,vjust =1),
        axis.title.x = element_blank(),
        axis.title.y = element_blank(),axis.text.y =element_blank(),
        text = element_text(family="serif"))+
  theme(plot.margin = unit(c(-0.5,0.15, 0.1,0), "cm"))+
  guides(color=guide_legend(ncol = 2))+
  theme(legend.key.size = unit(0.8, "lines"),
        legend.margin=margin(0,0,0,0),legend.box.margin=margin(-10,-10,-10,-10),
        legend.position = "right",
        legend.key = element_rect(color = NA, fill = NA),
        legend.spacing.x = unit(0.05,"cm"),legend.spacing.y = unit(-1, "cm"),
        legend.text=element_text(size=10))
legenda_perc_tur=get_plot_component(perc_tur, 'guide-box-right', return_all = TRUE)
perc_tur=perc_tur+theme(legend.position = "none")
```

7.2 Display panel Figure 2A

```r
cbp1 <- c("#0072B2","#e7298a","#fc4e2a","#bd0026","#88419d","#293352","#238b45",
          "#0072B2", "#E69F00", "#56B4E9","#FFDB6D", "#009E73", "#CC79A7",
          "#293352","#F0E442","#ff0000")

percentage$proc<-factor(percentage$proc,
                        c("Stochastic","Deterministic"))

perc_proc=ggplot(percentage,aes(x=group, y = perc.RC,colour=proc)) +
  geom_boxplot(width=0.5,outlier.colour =NA)+
  geom_jitter(position=position_dodge(0.5))+
  scale_color_manual(values = c(cbp1[c(8,4)]),name="")+
  labs(title="",y="Community assembly processes (%) ",x="Assembly process")+ylim(0,1.1)+
  theme_bw()+
  theme(panel.grid.minor = element_blank(),panel.grid.major = element_blank(),
        axis.text.x = element_text(size=10, color="black",angle =0,hjust=0.5,vjust =0),
        axis.title.x = element_blank(),
        text = element_text(family="serif"))+
  theme(plot.margin = unit(c(-0.5,0.15, 0.1,0), "cm"))+
  theme(legend.key.size = unit(1, "lines"),
        legend.margin=margin(0,1,0,1),legend.box.margin=margin(-10,-10,-10,-10),
        legend.position = "top",legend.box = "horizontal",
        legend.key = element_rect(color = NA, fill = NA),
        legend.spacing.x = unit(0.05,"cm"),legend.text=element_text(size=10))
legenda_perc_proc=get_plot_component(perc_proc, 'guide-box-top', return_all = TRUE)
perc_proc=perc_proc+theme(legend.position = "none")
```

7.3 Figure 2

Figure 2: Boxplots showing the percentage of variation in community organization accounted by different assembly mechanisms across the three spatial scales: local (i.e., individual mires), landscape (i.e., encompassing an entire national park), and regional (i.e., combining multiple national parks).



8 Analysis inside processes

```
#deterministic and stochastic
deter=data.frame(read_xlsx("data\\Assembly.xlsx",sheet="deter_scale",na=""))
stoch=data.frame(read_xlsx("data\\Assembly.xlsx",sheet="stoch_scale",na=""))

k<-kruskal.test(deter$perc.RC~deter$group)
k
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  deter$perc.RC by deter$group
## Kruskal-Wallis chi-squared = 9.8158, df = 2, p-value = 0.007388
```

```
kruskalmc(deter$perc.RC~deter$group)
```

```
## Multiple comparison test after Kruskal-Wallis
## alpha: 0.05
## Comparisons
##                      obs.dif critical.dif stat.signif
## Landscape-Local    16.116162     14.16553        TRUE
## Landscape-Regional  4.393939     17.16360       FALSE
## Local-Regional     11.722222     13.70598       FALSE
```

```r
k<-kruskal.test(stoch$perc.RC~stoch$group)
k
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  stoch$perc.RC by stoch$group
## Kruskal-Wallis chi-squared = 0.58084, df = 2, p-value = 0.7479
```

```r
kruskalmc(stoch$perc.RC~stoch$group)
```

```
## Multiple comparison test after Kruskal-Wallis
## alpha: 0.05
## Comparisons
##                     obs.dif critical.dif stat.signif
## Landscape-Local    2.287037     20.27829       FALSE
## Landscape-Regional 6.388889     22.99343       FALSE
## Local-Regional     4.101852     15.32895       FALSE
```

8.1 Deterministic processes

```r
#Heterogeneous selection and homogeneous selection
hete_selec=data.frame(read_xlsx("data\\Assembly.xlsx",sheet="he.sele_scale",na=""))
homo_selec=data.frame(read_xlsx("data\\Assembly.xlsx",sheet="ho.sele_scale",na=""))

k<-kruskal.test(hete_selec$perc.RC~hete_selec$group)
k
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  hete_selec$perc.RC by hete_selec$group
## Kruskal-Wallis chi-squared = 5.5079, df = 2, p-value = 0.06367
```

```r
kruskalmc(hete_selec$perc.RC~hete_selec$group)
```

```
## Multiple comparison test after Kruskal-Wallis
## alpha: 0.05
## Comparisons
##                     obs.dif critical.dif stat.signif
## Landscape-Local    4.805556    10.885615       FALSE
## Landscape-Regional 3.916667    12.711660       FALSE
## Local-Regional     8.722222     9.283284       FALSE
```

```r
k<-kruskal.test(homo_selec$perc.RC~homo_selec$group)
k
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  homo_selec$perc.RC by homo_selec$group
## Kruskal-Wallis chi-squared = 2.3123, df = 2, p-value = 0.3147
```

```r
kruskalmc(homo_selec$perc.RC~homo_selec$group)
```

```
## Multiple comparison test after Kruskal-Wallis
## alpha: 0.05
## Comparisons
##                     obs.dif critical.dif stat.signif
## Landscape-Local    5.416667    10.885615       FALSE
## Landscape-Regional 1.250000    12.711660       FALSE
## Local-Regional     4.166667     9.283284       FALSE
```

8.2 Stochastic processes

```
#Homogenizing dispersal, dispersal limitation and ecological drift
Homo_disp=data.frame(read_xlsx("data\\Assembly.xlsx",sheet="h.disp_scale",na=""))
Disp_limi=data.frame(read_xlsx("data\\Assembly.xlsx",sheet="d.limi_scale",na=""))
Eco_drift=data.frame(read_xlsx("data\\Assembly.xlsx",sheet="e.dri_scale",na=""))

k<-kruskal.test(Homo_disp$perc.RC~Homo_disp$group)
k
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  Homo_disp$perc.RC by Homo_disp$group
## Kruskal-Wallis chi-squared = 14.767, df = 2, p-value = 0.0006214
```

```
kruskalmc(Homo_disp$perc.RC~Homo_disp$group)
```

```
## Multiple comparison test after Kruskal-Wallis
## alpha: 0.05
## Comparisons
##                      obs.dif critical.dif stat.signif
## Landscape-Local     7.416667    10.885615       FALSE
## Landscape-Regional  7.083333    12.711660       FALSE
## Local-Regional     14.500000     9.283284        TRUE
```

```
k<-kruskal.test(Disp_limi$perc.RC~Disp_limi$group)
k
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  Disp_limi$perc.RC by Disp_limi$group
## Kruskal-Wallis chi-squared = 12.482, df = 2, p-value = 0.001948
```

```
kruskalmc(Disp_limi$perc.RC~Disp_limi$group)
```

```
## Multiple comparison test after Kruskal-Wallis
## alpha: 0.05
## Comparisons
##                      obs.dif critical.dif stat.signif
## Landscape-Local     6.694444    10.885615       FALSE
## Landscape-Regional  4.916667    12.711660       FALSE
## Local-Regional     11.611111     9.283284        TRUE
```

```
k<-kruskal.test(Eco_drift$perc.RC~Eco_drift$group)
k
```

```
##
##  Kruskal-Wallis rank sum test
##
```

```
## data:  Eco_drift$perc.RC by Eco_drift$group
## Kruskal-Wallis chi-squared = 7.7564, df = 2, p-value = 0.02069
```

```
kruskalmc(Eco_drift$perc.RC~Eco_drift$group)
```

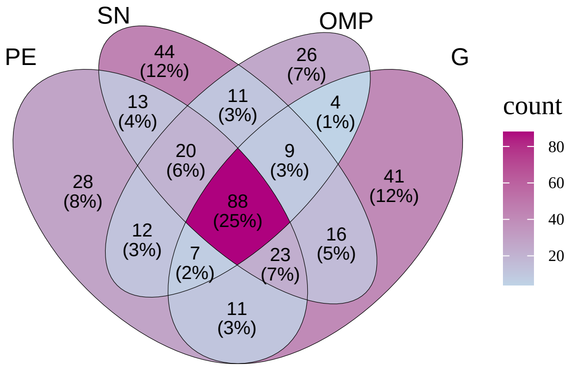
```
## Multiple comparison test after Kruskal-Wallis
## alpha: 0.05
## Comparisons
##                      obs.dif critical.dif stat.signif
## Landscape-Local     4.694444    10.885615       FALSE
## Landscape-Regional  5.916667    12.711660       FALSE
## Local-Regional     10.611111     9.283284        TRUE
```

9 Venn diagram

```
#Load dataframe
Abun_relative=as.data.frame(read_xlsx("data\\Assembly.xlsx",
                                      sheet="venn_diag_dataframe",na=""))
rownames(Abun_relative)=Abun_relative$Species

Abun_relative[Abun_relative>0]=1
Abun_relative=Abun_relative[,-1]
```

9.1 Display panel Figure S4

Figure S4: 354 diatom species were identified, representing 60 genera across the four national parks (Table S2). 25% of these species had widespread distribution at the regional scale and comprised almost 70% of the total relative abundance (Table S3). Taxa uniqueness varied across landscapes; Ordesa y Monte Perdido exhibited the lowest percentage of unique species (7%, n=26), followed by Picos de Europa (8%, n=28), Guadarrama (12%, n=41), and the Sierra Nevada (13%, n=45). In addition to showing low occurrences and being relatively uncommon, these unique species were also characterized by their low relative abundances (Table S3).



10 Traits

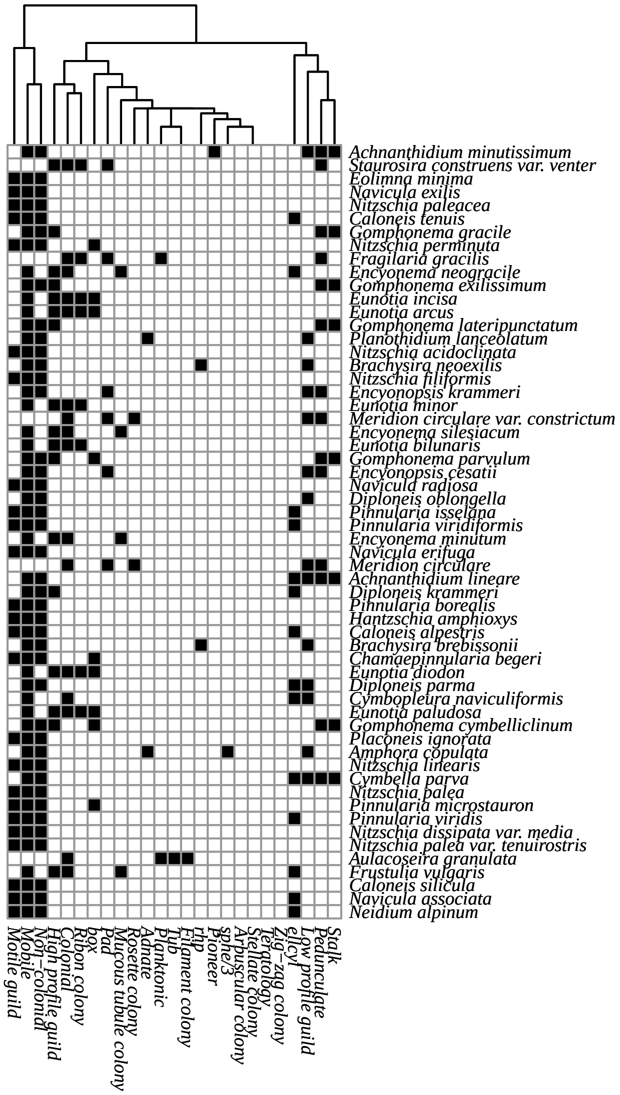
```
common_sp=as.data.frame(read_xlsx("data\\Traits.xlsx",sheet="traits_common",na=""))
unique_sp=as.data.frame(read_xlsx("data\\Traits.xlsx",sheet="traits_unique",na=""))

row.names(common_sp)<-common_sp$Species
row.names(unique_sp)<-unique_sp$Species

#Reorder by abundance
unique_sp<-unique_sp[order(unique_sp$Mean,decreasing = TRUE),]
common_sp<-common_sp[order(common_sp$Mean,decreasing = TRUE),]
```

10.0.1 Display panel Figure S5

Figure S5: Diatom traits in A) common and B) unique species in the regional scale, and C) the measure of sizes. In black boxes, the species presents these traits. The heatmap traits are grouped by cluster.

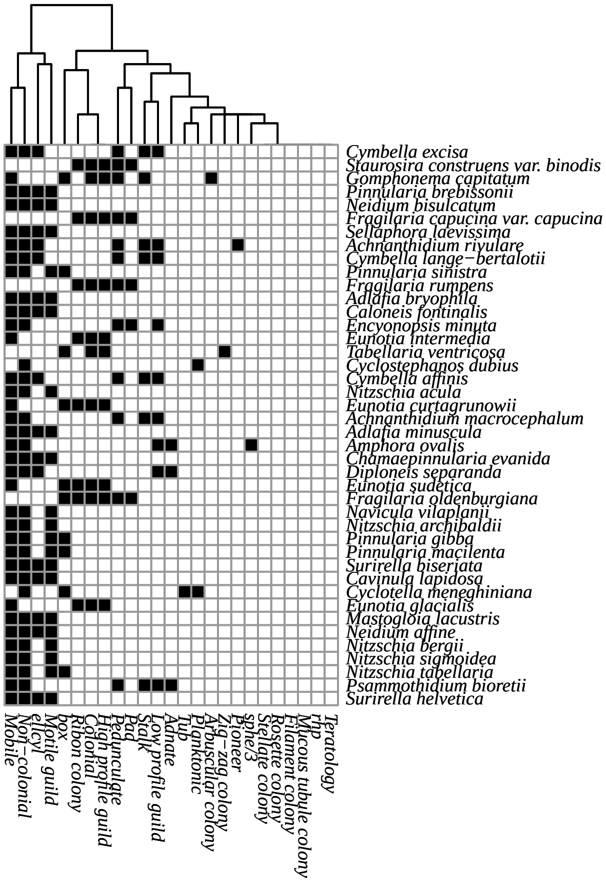
10.0.2 Display panel Figure S5 A

```
common=pheatmap(as.matrix(common_sp[,c(9:33)]),
                cluster_rows = FALSE,color=colorRampPalette(c("white","black"))(50),
                cellwidth = 5, cellheight = 5, fontsize = 7, legend = FALSE,
                fontfamily = "serif",
                fontface="italic")
common
```



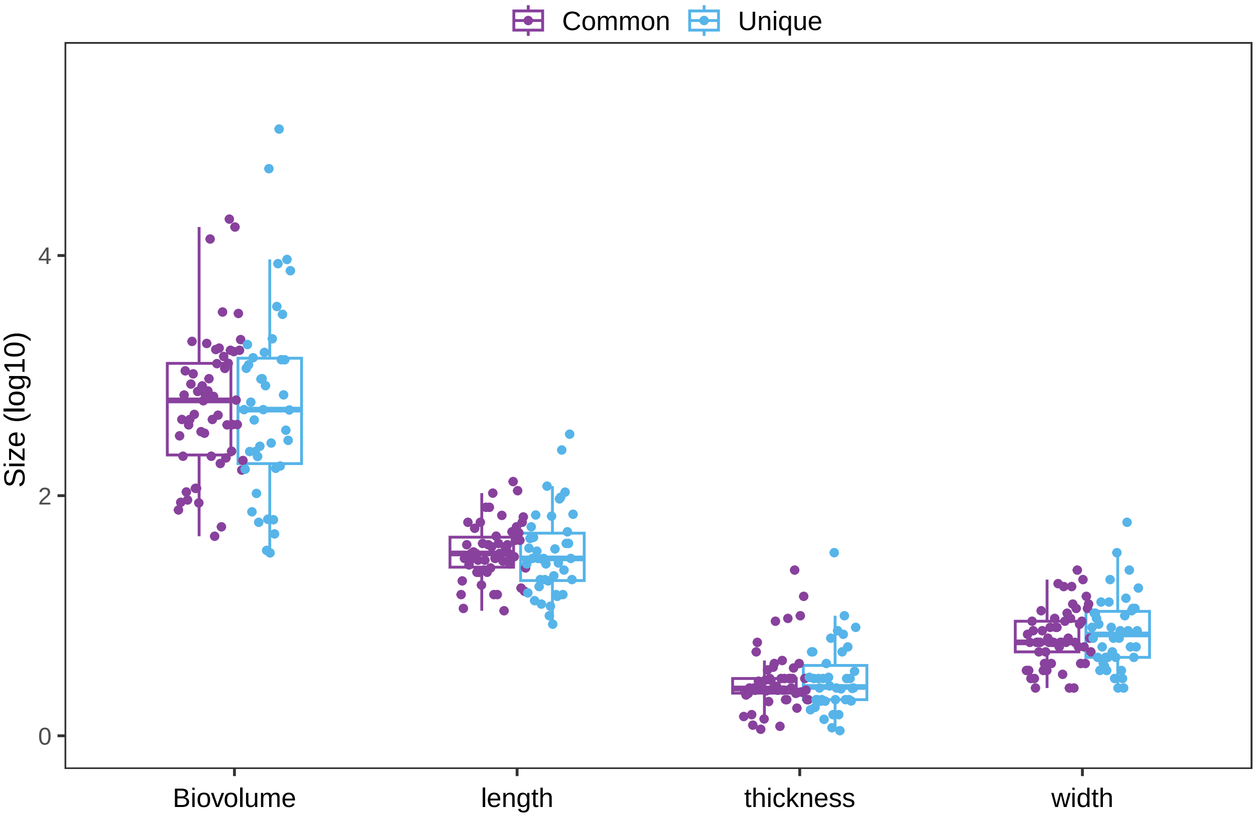
10.1 Display panel Figure S5 B

```
unique=pheatmap(as.matrix(unique_sp[,c(9:33)]),
                cluster_rows = FALSE, color=colorRampPalette(c("white","black"))(50),
                cellwidth = 5, cellheight = 5, fontsize = 7, legend = FALSE,
                fontfamily = "serif",
                fontface="italic")
unique
```



10.2 Display panel Figure S5 C

```
traits_size=as.data.frame(read_xlsx("data\\Traits.xlsx",sheet="Sizes",na=""))

cbp1 <- c("#bd0026","#88419d","#293352","#238b45","#0072B2", "#E69F00",
          "#56B4E9","#FFDB6D", "#009E73", "#CC79A7","#293352","#F0E442",
          "#ff0000")

sizes=ggplot(traits_size,aes(x=Category, y = log10(Size),colour=State)) +
  geom_boxplot(width=0.5,outlier.colour =NA)+
  geom_jitter(position=position_dodge2(0.4), size=1)+
  scale_color_manual(values = c(cbp1[c(2,7)]),name="")+
  labs(title="",y="Size (log10) ",x="")+ylim(0,5.5)+
  theme_bw()+
  theme(panel.grid.minor = element_blank(),panel.grid.major = element_blank(),
        axis.text.x = element_text(size=10, color="black",angle =0,hjust=0.5,vjust =0),
        axis.title.x = element_blank())+
  theme(plot.margin = unit(c(0,0, 0.5,0), "cm"))+
  theme(legend.key.size = unit(1, "lines"),
        legend.margin=margin(0,1,0,1),legend.box.margin=margin(-10,-10,-10,-10),
        legend.position = "top",legend.box = "horizontal",
        legend.key = element_rect(color = NA, fill = NA),
        legend.spacing.x = unit(0.05,"cm"),legend.text=element_text(size=10))
sizes
```



10.3 Normality test

```
tapply(traits_size$Size,traits_size$Category,shapiro.test)
```

```
## $Biovolume
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.22504, p-value < 2.2e-16
##
##
## $length
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.59722, p-value = 3.813e-15
##
##
## $thickness
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.46974, p-value < 2.2e-16
##
##
## $width
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.63585, p-value = 2.176e-14
```

```
bartlett.test(traits_size$Size~interaction(traits_size$State, traits_size$Category))
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  traits_size$Size by interaction(traits_size$State, traits_size$Category)
## Bartlett's K-squared = 3740.9, df = 7, p-value < 2.2e-16
```

10.4 Statistical test

```
wilcox.test(traits_size$Size[traits_size$Category=="Biovolume"]~
              traits_size$State[traits_size$Category=="Biovolume"])
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  traits_size$Size[traits_size$Category == "Biovolume"] by traits_size$State[traits_size$Catego
## W = 1246.5, p-value = 0.845
## alternative hypothesis: true location shift is not equal to 0
```

```
wilcox.test(traits_size$Size[traits_size$Category=="length"]~
              traits_size$State[traits_size$Category=="length"])
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  traits_size$Size[traits_size$Category == "length"] by traits_size$State[traits_size$Category =
## W = 1335, p-value = 0.4156
## alternative hypothesis: true location shift is not equal to 0
```

```
wilcox.test(traits_size$Size[traits_size$Category=="thickness"]~
              traits_size$State[traits_size$Category=="thickness"])
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  traits_size$Size[traits_size$Category == "thickness"] by traits_size$State[traits_size$Catego
## W = 1174, p-value = 0.7609
## alternative hypothesis: true location shift is not equal to 0
```

```
wilcox.test(traits_size$Size[traits_size$Category=="width"]~
              traits_size$State[traits_size$Category=="width"])
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  traits_size$Size[traits_size$Category == "width"] by traits_size$State[traits_size$Category ==
## W = 1146.5, p-value = 0.6195
## alternative hypothesis: true location shift is not equal to 0
```